LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

LLNL-TR-469231

LLNL Genomic Assessment: TMT Task 1.4 Final Report on Sequencing Knowledge Gaps

T. Slezak, M. Borucki, E. Vitalis, M. Torres, R. Lenhoff

February 7, 2011

Disclaimer

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

LLNL Genomic Assessment: TMT Task 1.4 Final Report on Sequencing Knowledge Gaps

LLNL-TR-469231

January 28, 2011

Correspondent:
Thomas R. Slezak, Associate Program Leader, Informatics
925-422-5746, slezak@llnl.gov

Contributing Authors:
Tom Slezak, 925-422-5746, slezak@llnl.gov
Monica Borucki 925-424-4251, borucki2@llnl.gov
Elizabeth Vitalis 925-422-0149, vitalis1@llnl.gov
Contributing Researchers:
Marisa Lam 925-423-2723, lam9@llnl.gov
Raymond Lenhoff 925-424-4034, lenhoff2@llnl.gov

Chemical and Biological Countermeasures Division
Global Security Program
Lawrence Livermore National Laboratory (LLNL)
Livermore, CA

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

Task 1.4 Final Report on Sequencing Knowledge Gaps

Under the DTRA Translational Medical Technologies (TMT) program Lawrence Livermore National Laboratory (LLNL) has been tasked with reviewing sequencing knowledge gaps for bacterial and viral pathogens of interest to TMT. The bacteria of interest are the biothreat agents *Bacillus anthracis, Brucella sp.*, *Burkholderia mallei* and *Burkholderia pseudomallei*, *Francisella tularensis* and *Yersinia pestis*. The primary initial viruses of interest were the viral hemorrhagic fever viruses, specifically Filoviruses (Ebola and Marburg), and Arena viruses (Guanarito, Junin, Lassa, Machupo, Sabia). During the course of this task it was determined that the viral list needed augmenting with other threat agents that are endemic in areas where DoD assets can be expected to be deployed. These include additional Arenaviruses (Chapare, Lujo), Flaviviruses (Tick-borne Encephalitis, Omsk, Kyasanur Forrest, Alkhurma, Japanese Encephalitis, Yellow Fever), Bunyaviruses (Crimean-Congo Hemorrhagic Fever, Rift Valley Fever), Heniparaviruses (Hendra, Nipah), and Alphaviruses (Venezuelan Equine Encephalitis, Easter Equine Encephalitis). These bacteria and viruses represent not only current human threats, but also future threats whether due to natural evolution or deliberate genetic alteration.

Executive Summary

TMT has made great progress in filling many major gaps in bacterial genome diversity. However, relatively little progress has been made in filling the much bigger gaps in viral genome diversity. At the present time no TMT performer has a demonstrated ability to perform large-scale production viral sequencing, something that is necessary if TMT is to make any progress in viral countermeasures. TMT has a current over-capacity for bacterial sequencing that needs to be ramped down to ensure that there are TMT-relevant reasons to sequence further bacterial isolates of several species (see below for details.) It is a critical need for TMT to establish a credible plan to be able to sequence hundreds of viral samples per year, including many deep quasi-species analyses. This plan ought to include establishing and/or maintaining relationships with scientists currently holding valuable strains or who would be predicted to be on the forefront of emerging outbreaks worldwide and able to obtain critical samples and/or sequence data.

TMT should add and test the ability to do bacterial metagenomic sequencing (e.g., not just a single isolated plaque but the entire population variation of complex clinical and environmental samples.)

Bacterial agents

1. Good progress has been made on filling bacterial diversity gaps, based on strain collections for which TMT (and other US) access is available. In general, we are at or nearing diminishing returns for sequencing pathogenic strains from the United States for *Bacillus anthracis* and *Yersinia pestis*, unless unusual phenotypical behavior is observed (e.g., related to speed of onset, transmission changes, high or low mortality, host shift, etc.) For these

organisms, there is higher potential value for sequencing avirulent (to humans at least) near-neighbor species to further comparative analyses.

2. It is less clear that we are nearing saturation for sequencing United States pathogenic strains of *Brucella*, *Burkholderia*, or *Francisella*. Continued judicious sequencing of available pathogenic strains of these species is still warranted (but see (3) below.) It is also certainly the case that we have only begun to sample the sequence diversity of the near-neighbors, particularly for *Burkholderia* and *Francisella* species. We note that these are extremely broad families and the virulence mechanisms have yet to be fully characterized.

3. Significant bacterial sequence gaps still remain for many bacterial agents in terms of diversity in other parts of the world. This is generally true for large portions of Asia, Africa, and the Middle East. The proposed global biosurveillance activities (known as CTR, Cooperative Threat Reduction or NLGC, Nunn-Lugar Global Cooperation) may someday provide an alternate path towards getting important strains sequenced in-country with subsequent data release for some locations, but the timeline on this is uncertain. TMT needs to monitor and participate in such activities, as continued access to international strains appears to be threatened.

4. The combination of increased bacterial pathogen sequence data and available technology for high-resolution genotyping (e.g., synthesized microarrays capable of identifying thousands of SNP states and the presence/absence of thousands of genes) indicate that in the absence of compelling phenotypic data, genotyping should be performed to determine if sequencing is warranted. There is little to be gained from continued re-sequencing of isolates that are nearly identical to ones already in the database. We note that DHS(Dept of Homeland Security) has launched a *Burkholderia pseudomallei* collaboration with Australia, involving TMT performers from LANL (Los Alamos Nat Lab) and LLNL, that will genotype ~100 Australian isolates using a SNP (single nucleotide polymorphisms) microarray to determine which ones warrant sequencing.

5. Resistant strains should continue to be sought and sequenced as a very high TMT priority. If the gene(s) conferring resistance is known to be carried on a movable element (e.g. plasmid), a cost effective strategy to consider would be sequencing of the plasmid only from numerous resistant strains to understand the evolution of the antibiotic resistance elements and sequencing the entire genome from a select few strains to identify potential chromosomal contributions.

Viral agents

1. In general, nearly all RNA viral pathogens have remaining sequence knowledge gaps in terms of known strain diversity:

a. Most existing genomes come from human hosts, illustrating a general lack of genomic sequence from strains carried by the natural reservoir and vector organism(s).

b. In many cases where these organisms are known, there are no samples whatsoever from large portions of the known natural ranges; indicating that a latent potential to infect humans in those areas may persist.

c. Often, genomic sequence knowledge of near-neighbor viruses that do not affect humans is minimal or absent. This leaves us unprepared when such viruses evolve to be capable of infecting humans (e.g., SARS).

2. Current RNA viral sequencing consists of creating a single consensus sequence. This collapses all the quasi-species information in a viral sample down to the highest-represented base at each position in the genome. To obtain the maximal useful information pertinent to broad-spectrum countermeasure design, TMT should strive to be at the forefront of changing the paradigm of viral sequencing to characterize the actual quasi-species variation.

3. Based on currently available viral sequence and current knowledge about likely TMT access to international strains containing the types of viral diversity outlined above, there are real reasons for concern as to whether TMT can actually obtain the viral sequence information needed to successfully drive anti-viral research improvements. Possible ways to overcome this include:

    a. Leveraging existing CDC (Centers for Disease Control, Atlanta) and UTMB (University of Texas Medical Branch) connections to obtain either access to strains or to get those strains sequenced by others. Note that the current WHO (World Health Organization) connections, while an invaluable source of potential novel viruses, seem less likely to be able to provide the broader spectrum of known and unknown circulating diversity.

    b. Sponsoring viral diversity research, including sequencing, in countries where strains are endemic. This might be accomplished via the DoD GEIS (Dept of Defense Global Emerging Infections Surveillance) labs, in some regions, or else via the CTR/NLGC efforts as they materialize.

4. TMT needs to put as much emphasis on, and energy into, filling viral diversity knowledge gaps as it has on bacterial sequencing. This is vital for TMT to make progress on anti-viral countermeasures.

We note that the following summaries were derived from the detailed reports we submitted over the course of the last 18 months and were based on the sequence status at the time the analyses were performed.

Bacterial Agent Summary

Most of the organisms sequenced to date under TMT funding have been bacteria from the major threat agent category or their near-neighbors. Some specific comments on each of these major threat agents are given below. A more general cautionary comment is that it appears that we are rapidly approaching a point where major geographical sequence knowledge gaps simply cannot be filled because access to samples is not feasible. Efforts should be redoubled to ensure that TMT-funded sequencing capacity is being focused on strains relevant to MCM (Medical CounterMeasures) design, rather than general evolutionary diversity studies. In general, these would include close (avirulent) near-neighbors and any strains showing unusual phenotypes (e.g., very high or low mortality, unusual resistance, host range change, etc.) Recent TMT-funded sequencing of strains (from multiple species) containing the NDM-1 resistance plasmid are a good example of appropriate focusing.

It should be anticipated that TMT-funded bacterial sequencing will rather quickly ramp down to occasional new near-neighbor strains and MCM-relevant strains, particularly strains with broad antibiotic resistance, and perhaps with an occasional burst of geographical gap-filling if and when strains become available. It should be clearly noted that the bulk of TMT sequencing should rapidly evolve to be viral strains, from the standpoint of having the highest total impact on overall TMT broad-spectrum countermeasure goals.

Bacillus anthracis

The spore stability of this species keeps it near the top of the bacterial agents despite an abundance of existing sequence data and reduced antibiotic resistance mechanisms when compared to gram negative bacteria. Due to lack of solid correlation between a particular genotype and human pathogenicity, studies of pathogenicity have focused on the plasmids that confer virulence such as the capsule and edema factor, lethal factor and protective antigen.

Priority Queue 1

- To better understand virulence, additional isolates *of B. cereus* obtained from human inhalational anthrax cases should be sequenced.

Unusually virulent strains or alternatively, avirulent and/or vaccine strains ought to be sequenced if they should become available (ex vaccine strains: *B. anthracis Carbosap* and *B. anthracis strain* 55 (see input from National Veterinary Services Laboratories (NVSL))).

Priority Queue 2

- Additional *B. anthracis* isolates from Africa should be sequenced.
- Additional *B. anthracis* isolates from N. America that comprise the C group should be sequenced.
- Near neighbor sequence is needed, especially, *B. mycoides and B. weihenstephanensis.*

Yersinia pestis

Given the extensive existing data for this species, and the information from Northern Arizona University (NAU) that 160 new genomes from China are being sequenced, and will be publicly released in the near future, it has a lower priority in general for TMT's sequencing program than most other bacteria.

Priority Queue 1

- To better understand virulence, isolates from *Y. pestis Pestoides* and *Y. pestis* biovar *Microtus* genomes (generally avirulent to humans but virulent in rodents) should be sequenced.
- Very little is known about antibiotic resistance in this bacteria, and therefore, should any resistant strains be discovered, these ought to be top priority for sequencing.
- Highly virulent strains and avirulent strains (if they can be obtained) from the ORI molecular group may provide valuable information about virulence in *Y. pestis.*

Priority Queue 2

- Isolates of *Y. pestis* that are similar to *Y. pestis Angola* should be sequenced due to their diversity.
- To broaden the scope of near neighbors, *Y. frederickensii*, *Y. rohdei,* and *Y. ruckeri* and others should be sequenced.

Francisella tularensis

Unknown environmental reservoirs of this bacteria and poorly defined mechanisms of virulence make this a priority for additional focused genome sequencing.

Priority Queue 1

- *F. tularensis* subsp. *mediasiatica* (*F.t. mediasiatica*) virulence is uncertain due to the scarcity of isolates to date. Whole genomic sequencing is needed for the subspecies *F.t. mediasiatica*, particularly from Japan.
- Any avirulent and/or vaccine strains of *F. tularensis* that may be discovered (in addition to the LVS strain) should be top priority to help understand virulence mechanisms in these bacteria.
- Any strains exhibiting antibiotic resistance ought to be sequenced, as very little is known about potential antibiotic resistance mechanisms in these bacteria.

Priority Queue 2

- Additional *F.t. holartica* isolates from Japan should be sequenced, and *F.t. holartica* isolates from California subclade B.br.002/003 should be sequenced.
- *F.t. novicida* isolates from Australia that are known to cause human infection should be sequenced.
- Near neighbors *F. pisicicida* and species in the *Wolbachia* genus should be sequenced. (We note that NBACC (National Biodefense Analysis and Countermeasures Center) has

draft sequenced *F. pisicicida*, but it has not been shared with DTRA or made public yet, to our knowledge.)

***Brucella* species**

To date, *Brucella* has lacked acute extreme virulence, and thus is our lowest priority for Category A bacterial sequencing for TMT's needs.

Priority Queue 1

- Little is known about its virulence or antibiotic resistance mechanisms, particularly due to the lack of strains sequenced from human infections. There are strains showing antibiotic resistance (for ex*ample B. abortus* strain RB51), and these should be top priority for this organism in addition to pathogenic strains isolated from humans.

Priority Queue 2

- Additional human isolates of *Brucella melitensis*, *B. suis* and *B. abortus* should be sequenced.
- Additional human isolates from newly discovered *Brucella* species, *B. pinipedialis, B. ceti*, and *B. inopinata* should be sequenced.
- For complete phylogenetic data, isolates from biovar 5 of *B. abortus* as well as biovars 1, 2 and 4 of *B. suis* should have at least one complete genome sequenced.
- *Brucella* from the Middle East (especially *B. melitensis*) should be sequenced if available.

***Burkholderia* species**

A large amount of unknown genetic variation for *B. psuedomallei* due to newly discovered pathogenicity island differences combined with multiple mechanisms of antibiotic resistance make sequencing additional genomes of high importance.

B. mallei is a human pathogen of poorly understood virulence due to the scarcity of natural human infections. In addition, only 11 complete genome sequences of *B. mallei* exist, four of which are derived from a single infection.

Priority Queue 1

- There is a need for genomes of isolates of *B. pseudomallei* from non-human hosts.
- Additional isolates from Thailand (source of isolates associated with human disease) are needed to better define the spectrum of *B. pseudomallei* human virulence.
- There are currently no avirulent or vaccine strains of either *B. pseudomallei* or *B. mallei*, and therefore, if any should be discovered, they ought to be top priority to help understand virulence mechanisms in these bacteria.

Priority Queue 2

- Additional *B. mallei* genomes should be sequenced, especially from S. America and Africa.

- Genomic sequence from near neighbors other than *B. cepacia* is needed.

DHS has initiated an international collaboration with Australia. They will require assistance sequencing at least 50 isolates of *B. pseudomallei* over the next 2 years. The plan is to prioritize ~100 isolates for sequencing using a *B. pseudomallei* SNP microarray customized by LLNL. NBFAC (National Biodefence Analysis Center) plans to sequence a few of the isolates. Both clinical and environmental isolates will be sequenced. DHS has engaged LANL to perform the bulk of the sequencing. It is presumed that these genomes will be freely shared with TMT.

Viral Agent Summary

Filoviruses

Ebolavirus: Whole genome sequence data is needed from human and animal hosts and vectors. Currently, there is no whole genome sequence data available for 8 of the 18 Ebolavirus outbreaks or from infected (viral RNA positive) bat and ape samples. Ecological studies suggest that Ebolavirus is present in countries that have not yet reported cases of Ebolavirus hemorrhagic fever (such as Cameroon and Ghana); potentially infected samples from this region should be included in the study if possible. Potential sources of this material include CDC, UTMB, Wildlife Conservation Society, and NIAID.

Marburgvirus: Whole genome sequence data is available from human isolates from all of the Marburgvirus epidemics and from one species of bat that serves as a Marburgvirus vector. Ecological modeling studies and evidence of Marburgvirus infection in bats indicate that Marburgvirus is present in countries that have not yet reported cases of Marburg hemorrhagic fever; potentially infected samples from this region should be included in the study if possible. Strains of Marburgviruses from the same outbreak may exhibit high overall genetic diversity, therefore multiple isolates should be sequenced from each outbreak. Genetic analysis indicate that genetic diversity is not correlated with geographic distance, therefore multiple samples from the same site and outbreak should be examined for genetic diversity. Potential sources of this material include CDC, UTMB, Wildlife Conservation Society, and NIAID (Nat Institute of Allergy and Infectious Diseases).

(Note: These reports were provided to Stuart Nichol at CDC and were used by him to prepare his TMT proposal for sequencing 72 Ebolavirus and Marburgvirus strains, to help fill the described gaps. Sequencing is now underway at CDC.)

Arenaviruses

Eight new arenavirus species have been discovered in the last five years, three of which cause hemorrhagic fever in humans. One of the newly detected hemorrhagic fever arenavirus species, Lujo virus, was discovered in South Africa in 2008 and had a mortality rate of 80%. It is likely

that additional genetically distinct pathogenic arenaviruses remain to be discovered in Africa and elsewhere.

Lassa virus: Lassa virus is the most common imported viral hemorrhagic fever agent in Europe and the United States. Lassa virus strains are antigenically and genetically diverse and may differ in nucleotide sequence by up to 27%. Currently ten Lassa virus strains have been sequenced, however, critical data such as date of isolation, host, and location are not available for many of these strains, and all may be human isolates. Therefore, it is important to sequence a comprehensive and diverse set of human and rodent isolates in order to design assays capable of detecting genetically heterologous Lassa virus strains associated with human disease. This has been recently illustrated by the recent discovery of a new Lassa virus strain that appears to represent a new lineage within the species.

Junin virus (Argentine hemorrhagic fever): A large number of Junin virus strains have been isolated from humans and rodents, but only four strains have been fully sequenced, all of which are laboratory passaged strains obtained from humans decades ago. Additionally, three of the four strains sequenced are from derived from the same isolate. Analysis of partial genome sequence data of 39 strains of Junin virus isolates obtained from human and rodent hosts indicates that there is up to 13% nucleotide differences between strains. Whole genome analysis is needed for viral strains isolated more recently from humans and from rodent vectors from different geographical regions of the endemic area. Isolates from both mild and severe cases of AHF should be sequenced as should strains of near neighbors, Machupo virus and Tacaribe virus. Potential sources of strains include Mike Bowen and Stuart Nichol, (CDC); Steve St. Jeor (Univ. of Nevada, Reno); and Bob Tesh and Tom Ksiazek (UTMB).

Guanarito virus (Venezuelan hemorrhagic fever): The only fully sequenced genome of Guanarito virus is an isolate from a human case of Venezuelan hemorrhagic fever that occurred in 1990, therefore more recent isolates from humans and rodent vectors need to be sequenced. Isolates of near neighbors Amapari, Cupixi, Tacaribe, and Sabia viruses should be sequenced as one completely sequenced genome exists for each of these viruses. Sources of viral isolates include CDC (Bowen, Nichols) and UTMB (Fulhorst, Weaver, Tesh), however, in the case of near neighbor species, it is likely that no additional virus isolates are available.

Machupo virus (Bolivian hemorrhagic fever): Whole genome sequence data exists for two strains of Machupo virus. Both sequenced strains are from human cases and were isolated decades ago (1963 and 1971). Phylogenetic analysis of the N protein gene sequence data from human and rodent strains obtained between 1963-2000, identified eight distinct genetic lineages within the Machupo species with up to 13% nucleotide difference between strains. Suggested strains for whole genome sequence analysis include isolates obtained from the rodent vector (vesper mouse) as well as isolated obtained from more recent human cases. Potential sources of material include CDC (Nichol), UTMB (Fulhorst, Ksiazek, Weaver, Tesh, Peters), and Univ. of New Mexico (Salazar-Bravo, Yates).

Hemorrhagic fever-associated arenaviruses for which the animal vector is unknown: Sabia virus, Chapare virus, and Lujo virus: Three arenavirus species (Sabia virus, Chapare virus, and Lujo virus) have been isolated from human hemorrhagic fever cases but have not yet been detected in an animal vector, therefore the geographical range and ecology of these viruses remains undetermined. In each case a single complete genome sequence is available because either only one clinical isolate was available (Sabia virus and Chapare virus), or multiple samples obtained from different cases from a single outbreak were determined to be genetically identical. Provided follow up studies are done on potential rodent vectors and/or additional human cases occur, future sources of isolates may include Stuart Nichol (CDC) and the World Reference Center of Emerging Viruses and Arboviruses (UTMB).

Flaviviruses

Tick-borne encephalitis (TBE) virus: The incidence of TBE has increased dramatically with a 400% increase in morbidity occurring in Europe from 1997 to 2003. Additionally, TBE virus appears to be spreading in range and infecting new species of ticks. Whole genome sequence data is available for 24 strains from six countries. Most sequenced TBE virus strains were obtained from humans and were collected in Russia, therefore temporally and geographically diverse isolates from a variety of tick species and animal and human hosts should be prioritized for sequencing. Phylogenetic analysis indicates that louping ill virus is a subtype of TBE virus, and although it is primarily a veterinary pathogen, it has been associated with a variety of disease syndromes in humans including encephalitis and hemorrhagic fever. Although a variety of isolates exist, only one strain of louping ill virus has been fully sequenced. Potential sources of strains of TBE virus and near neighbors include Michael Holbrook (UTMB) and Alexander Pletnev (Laboratory of Infectious Diseases, NIAID).**Tick-borne hemorrhagic fever flaviviruses - Omsk, Kyasanur Forest, and Alkhurma:** Three members of the tick-borne encephalitis serocomplex are known to cause hemorrhagic fever in humans: Omsk hemorrhagic fever virus (OHFV), Kyasanur Forest disease virus (KFDV), and Alkhurma virus (ALKV), which is a subtype of KFDV. Compared to OHFV, KFDV has the highest mortality rate and occurs much more frequently. Although many KFDV isolates exist and hundreds of human cases occur annually, only two genome sequences are available for KFDV, one is the reference strain isolated in 1957, and the second has no metadata available. Although ALKV may infect a variety of hosts, only one strain (human) of ALKV has been fully sequenced. Finally, three strains of OHFV have been sequenced, all of which were isolated many decades ago. Therefore, a combined total of six whole genome sequences are available from this group of viruses. Although available sequence data from OHFV, KFDV, and ALKV indicate that the genomes of these viruses are stable, cases of KFD continue to occur frequently despite routine vaccination and this may be due to antigenic drift. Comprehensive genetic characterization of currently circulating strains of KFDV is required for the development of improved vaccines and rapid diagnostic tests. Therefore, any and all representatives of this group of virus as well as near neighbor viruses should be prioritized for sequencing. Potential sources of strains of KFDV,

ALKV, and OHFV virus and near neighbors include Michael Holbrook (UTMB) and Stuart Nichol (CDC).

Japanese Encephalitis Virus: Japanese encephalitis is a mosquito-borne viral infection that is prevalent in large parts of Asia and New Guinea. The virus has extended its geographic range both east and west from the region in Asia where it was first identified to India and Australia. There are 73 whole genome sequences for this virus with about half of the sequenced isolates from China as well as isolates from Japan, Korea, India, Thailand, and Australia. More genome sequences are needed for isolates from Malaysia, and Indonesia, where all five genotypes of this virus, and possibly the ancestral origin of this virus, is thought to exist Sequences from New Guinea and Pakistan where this virus is spreading are also lacking. No complete genome sequences of isolates from the known avian hosts have been identified, however this may be due to the nonclinical nature of the infection in this host. Potential sources of JEV are Alan Barrett and Bob Tesh (UTMB), Bruce Innis (Walter Reed), Barbara Johnson or Lyle Peterson (Ft. Collins branch the CDC), and Tom Solomon (University of Liverpool).

Yellow fever virus: Yellow fever is caused by a mosquito-borne virus that can cause limited febrile illness as well as jaundice and hemorrhagic disease in humans. Currently 90% of the 200,000 cases per year that result in 30,000 deaths occur in Africa despite the availability of a good vaccine. There are five African and two South American genotypes of the virus. Complete genome sequence is available for a twenty genomes however half of the complete genome sequences are closely related to the 17D vaccine strain. While East Africa is thought to be the origin of yellow fever virus (YFV) and hence contains the most genetically diverse isolates (, samples of virus from East and Central Africa are scarce and more samples ought to be collected and genomes sequenced from these regions. Complete genomic sequences are particularly scarce in the sequence database for the two South American genotypes. In addition, only one complete viral genome sequence for YFV isolated from non-human primates exists in the Genbank sequence database. As this report was being assembled, information was published by UTMB and Brazilian scientists that helps meet these gaps, and has been incorporated. This update to the report underlines the dynamic nature of characterizing sequence gaps and the need to keep informed of ongoing sequence efforts worldwide. Just after submission of the report, and undiagnosed HV outbreak in Uganda was determined to be caused by YFV. Samples from this outbreak would be top priority if available. Potential sources of well characterized YFV include Alan Barrett and Bob Tesh (UTMB), and Manfred Weidmann (Institut Pasteur de Dakar).

Bunyaviruses

Crimean-Congo hemorrhagic fever (CCHF) virus: Whole genome sequence data is available for 22 isolates of CCHF virus. Most sequenced strains were obtained from humans and were collected more than a decade ago. The genome of CCHF virus is very plastic due to point mutations, segment re-assortment and possibly recombination. Genetic diversity is generally correlated with geographic location and possibly tick species rather than date of isolation, source

of infection, or virus strain virulence. Therefore, temporally and geographically diverse isolates from a variety of tick species and animal hosts should be prioritized for sequencing. Additional temporally and geographically diverse human isolates should also be sequenced. For example samples and sequence ought to be obtained from a recent CCHF out break in India, the first documented cases in India. There is limited or no full genome sequence data available for CCHF virus near neighbors Hazara virus, Nairobi sheep disease virus, Dugbe virus, and Kupe virus; more isolates should be sequenced if available. Potential sources of strains include Stuart Nichol, Varough Deyde, and Barry Miller (CDC), and Connie Schmaljohn (USAMRIID), and Tom Ksiazek (UTMB).

Rift Valley fever virus (RVFV): RVFV is a mosquito-borne pathogen that is endemic throughout most of Africa and the Arabian Peninsula. Analysis of whole genome sequence data of temporally and geographically diverse strains of RVFV isolates indicates that there is about 5% nucleotide difference between strains, and multiple strains may circulate within a single outbreak. Evidence of naturally occurring genome segment assortment has been documented. Whole genome sequence is available for over 50 strains of RVFV, however, no isolates from the last five major epidemics have been sequenced. Indeed, over the last 10 years thousands of human RVF cases have been reported, resulting in hundreds of deaths and only one human isolate from this period has been fully sequenced. Similarly, almost no near neighbor species have been sequenced. Indeed, Punta Toro virus, which is often used in place of RVFV to model RVF disease in rodents, has yet to be sequenced. Potential sources of strains include Stuart Nichol (CDC) and Bob Tesh (UTMB).

Henipaviruses

Henipaviruses: Hendra virus and Nipah virus are currently the sole members of the Henipavirus genus, which is within the *Paramyxaviridae* family, a family of viruses which includes a long list of human and veterinary pathogens. To date, a total of nine Nipah virus genomes from three outbreaks have been sequenced, therefore, representatives of the other eleven Nipah virus outbreaks need to be sequenced. There is only one Hendra virus isolate that has been sequenced and no metadata is available for this strain. Recently, viral RNA was isolated from African fruit bats (*Eidolon helvum*) that were similar in sequence to that of the Henipaviruses. This sequence data along with seroprevalence data indicate that henipaviruses are present in this abundant species of African bats and this potentially greatly extends the range of these viruses Although no virus was isolated and only part of the viral genome was amplified from bat tissues due to low virus titer, continued efforts should be made to identify new members of the Henipavirus genus and to characterize the viruses genetically and phenotypically. Potential sources of strains of Nipah and Hendra virus strains and near neighbors include Paul Rota and William Bellini from CDC (Measles, Mumps, Rubella, and Herpesvirus Laboratory Branch), Bryan Eaton, Lin-Fa

Wang and Kim Halpin from CSIRO (Australian Animal Health Laboratory), and Jonathan Epstein at the Consortium for Conservation Medicine, New York.

Alphaviruses

Venezuelan equine encephalitis virus (VEEV): VEEV is a mosquito-borne virus capable of causing large outbreaks of encephalitis in humans and horses. VEE outbreaks occur sporadically and thirteen subtypes of VEEV exist, however only two are associated with large epidemics. Although one epidemic was likely caused by the presence of viable virus in formalin inactivated vaccine, epidemic strains may also emerge from endemic strains that have acquired a pattern of mutations that cause a shift in the virus transmission cycle thus enabling the virus to infect new mosquito vectors and vertebrate hosts (horses and humans). VEEV is a member of the VEE antigenic complex, a group of closely related viruses that includes six serotypes (I-VI) and 13 serologically distinct subtypes. Whole genome sequence is available for 32 isolates belonging to the VEEV complex, 25 of which are isolates of VEEV (serotype I). Additional isolates of serotypes II-VI need to be sequenced. Because experimental data indicate that minor genetic variations can cause enzootic viruses to emerge as enzootic/epidemic viruses and the role of viral population variation (quasispecies) in emergence of outbreaks has not been defined, the use of ultra-deep sequencing methods should be employed to address the dynamics of epidemic subtype emergence. Potential sources of strains of VEE virus strains and near neighbors include Scott Weaver and Bob Tesh (UTMB), Ann Powers, Aaron Brault and Richard Kinney (CDC, Ft. Collins, CO), and Patricia Aguilar and Tadeusz Kochel (Naval Medical Research Center Detachment).

Eastern equine encephalitis virus (EEEV): EEEV is a mosquito-borne virus capable of causing large outbreaks of encephalitis in humans and horses. In North America, EEEV infection has a very high mortality rate in humans, and survivors often suffer severe neurological sequelae. Interestingly, EEEV infections from South American isolates are generally subclinical. Although EEEV is divided into two antigenic varieties and four lineages, only eleven isolates have been sequenced and eight of these are from the North American variety (Lineage I). Most sequenced strains were collected from mosquitoes and only one human isolate has been sequenced. EEEV isolates exist from a variety of hosts, vectors, years, and geographical locations and efforts should focus on sequencing strains that represent this diversity. Potential sources of strains of EEEV strains and near neighbors include Scott Weaver and Bob Tesh (UTMB), Mike Turell (USAMRIID), and Aaron Brault (CDC, Ft. Collins).

References

See the individual reports for detailed references.